Supplementary Method

Participants

Participants were sixty-seven undergraduate students (41 female) at the University of Notre Dame (mean age 19.6, SD=1.4; range 18-25). They were recruited through psychology courses or flyers posted around campus and received compensation in the form of course credit or cash payment. Exclusion criteria included self-reported mental illness (e.g. major depression, anxiety disorder) and use of any medication affecting the central nervous system. Participants were instructed to refrain from eating and drinking for an hour prior to and during the study. They were also required to refrain from intake of caffeine, alcohol, or tobacco during the 24 hours prior to the start of the study and throughout the study. All participants gave written informed consent and the standing ethics committee at the University of Notre Dame approved test protocols.

Materials

Trier Social Stress Test

Stress was induced using the Trier Social Stress Test (TSST; Kirschbaum et al., 1993), a procedure that reliably elicits moderate psychosocial stress in laboratory settings. The task consisted of an anticipation period of written speech preparation (10 min) and a delivery period that required participants to deliver a free speech from memory (5 min) and perform mental arithmetic (5 min) in front of two "judges," who were trained experimental confederates. Participants' notes made during the preparation phase were abruptly taken from them immediately before the delivery period and they were asked to give the speech from memory. Furthermore, participants were told they were being recorded for vocal and video analysis with the use of a microphone and a video camera that was connected to a 19-inch TV screen. To make the task additionally stressful, participants could watch themselves deliver the tasks in this

TV monitor. The arithmetic task consisted of participants subtracting 13 from 1022 serially and without stopping for 5 minutes; if mistaken, they had to start again from 1022.

In contrast, participants in the control group performed the 10-min delivery period sitting in an empty room with no microphone, no video camera, and no judges, but still delivering the speech with their initial notes in hand. Participants were told of their group assignment before the "speech" preparation, to help avoid anticipatory stress.

Deese-Roediger-McDermott (DRM) Memory Task

Participants listened to 15 lists (Stadler et al., 1999) of semantically related, neutral words (12 words per list) through headphones while sitting in front of a computer. Words were presented at a rate of one word per second and at the final word of each list there was a 12-second break, followed by a one-second tone, followed by two seconds of silence, and then the start of the next list. Participants were asked to pay attention to the words because they were going to be tested in the next session. Twenty-four hours later, participants returned to the lab for the testing session, where a free recall test was given, followed immediately by a computerized recognition test. The free recall test was a simple pen-and-paper version where participants were instructed to write down all the words they could remember from the last session.

Salivary Cortisol Analysis

Salivary cortisol was measured in response to the TSST as a measure of hypothalamic-pituitary-adrenal (HPA) axis activity to stress. Six saliva samples (see *Design and Procedure* and Figure S1) were collected from each participant in plastic tubes using the passive drool method, in which participants directly expectorate into a tube (Shirtcliff et al., 2001).

Saliva samples were assayed in duplicate to determine cortisol levels using a highly sensitive enzyme immunoassay (Salimetrics, State College, PA). The test used 25 µL of saliva per determination, has a lower limit of sensitivity of .003 µg/dL, standard curve range from .012

µg/dL to 3.0 µg/dL, an average intra-assay coefficient of variation of 3.5% and an average inter-assay coefficient of variation of 5.1%.

Design and Procedure

Participants arrived at the laboratory in the afternoon between 3-4pm and signed informed consent. After 20 minutes had passed to allow acclimatization to the laboratory environment, they were asked to provide a 5mL sample of their saliva (T_0) in order to establish a baseline level of cortisol. After participants were finished with the first saliva sample, they listened to the 15 DRM lists. Immediately following encoding, participants were randomly assigned into a stress group (n=33, 18 females) or a control group (n=34; 23 females) and either performed the TSST or control task, respectively. Directly after the TSST or the control task participants provided a second saliva sample (T_1), followed by a third saliva sample 20 min later (T_2), and a fourth sample an hour after the conclusion of the TSST (T_3). After the completion of the fourth saliva sample and the questionnaires, participants left the laboratory to continue with their everyday activities.

Twenty-four hours later, participants returned for the memory retrieval tests. Upon arrival, they provided a fifth saliva sample (T_4). After the completion of this saliva sample participants were given a free recall test followed by a recognition test to fully probe memory, always in this order (e.g., Payne et al. 2006, 2007). Immediately after, a sixth and final saliva sample was provided (T_5), and the participants were debriefed and thanked for their participation. For a timeline of the experiment see Figure S1.

Cortisol and Memory Measures

Cortisol data is presented and analyzed in nmol/L (Supp. Fig. S2). Area under the curve with respect to increase (AUCi) was calculated to quantify change in cortisol concentration across time with respect to individual baseline cortisol concentration. AUCi was calculated as:

$$AUCi = \left(\sum_{i=1}^{n-1} \frac{(m_{(i+1)} + m_i) \cdot t_i}{2} \right) - \left(m_1 \cdot \sum_{i=1}^{n-1} t_i \right)$$

with t_i denoting the individual time interval between the two measurements, m_i denoting the cortisol concentration at each given time, m_1 denoting the cortisol concentration of the initial baseline sample and *n* referencing the total number of measurements (see Pruessner et al., 2003). In addition to using this measure as a task manipulation check between the stress and control groups, the stress group was sorted based on calculated AUCi and were divided at the midpoint to create a median split between low and high stress responders.

Memory was assessed using both recognition (raw and corrected) and free recall testing. For the recognition test, true rate was defined as old-hit (old response to an old item) and false rate as related-false alarm (old response to a critical lure). Foil rate was considered as an old response to a new, unrelated word. Our main recognition measures were corrected true recognition, calculated by subtracting foil rate from true rate and corrected false recognition calculated by subtracting foil rate from false rate (e.g., Diekelmann et al., 2008). Additionally, we used the bias measure B_r from two-high-threshold theory (Snodgrass & Corwin, 1988; Feenan & Snodgrass, 1990). True B_r was calculated with the following formula: foil rate/1-(hit rate-foil rate). False B_r was calculated with the following formula: foil rate/1-(false rate-foil rate).

Supplementary Results

Cortisol analysis for high cortisol responders

In order to examine memory performance in both the stress group as a whole, and in subjects who mounted a cortisol response to the TSST, we split the stress group into low and high (n=16) cortisol responder groups. We created these groups by applying a median split using our cortisol change measure AUCi (see *Design and Procedure* section). Although averaged cortisol levels were elevated by the stress manipulation in all subjects in the stress

condition (see Supp. Fig. S2A), this pattern became much stronger when assessing high cortisol responders and controls separately (see Supp. Fig. S2B). A 2 (Group: control vs. high responders) x 6 (Time) mixed ANOVA with Greenhouse-Geisser correction was performed on the cortisol data, with Time as a repeated factor. Both main effects of Time ($F=3.79$, $p=.007$) and Group were significant ($F=5.71$, $p=.02$). More importantly, there was a significant Group x Time interaction, $F(6,198)=18.10$, $p<.001$. Bonferroni-corrected comparisons showed that high responders, compared to controls, had equal cortisol levels at baseline ($p=.86$) and significantly higher salivary cortisol at time points T_1 ($p<.001$) and T_2 ($p<.001$) and T_3 ($p=.04$). No significant differences were found between the groups during the second day (T_4, T_5), all p's >.79.

Raw recognition

True recognition rate was calculated by dividing the "old" responses given to a study word by the total number (45) of presented study words (hit rate). False recognition rate was calculated by dividing the "old" responses given to a critical lure by the total number (15) of critical lures presented (false alarm rate to critical lures). Foil recognition rate was calculated by dividing the "old" responses to unrelated foils by the total number (32) of unrelated foils presented (false alarm rate to unrelated foils); See Supp. Table S1 for a breakdown of recognition memory performance by group. A 2 (Group: stress vs. control) x 3 (Word Type: True vs. False vs. Foil) mixed ANOVA, with Word Type as repeated factor, revealed significant main effects of Group, $F(1, 65)=4.89$, $p=.03$, and Word Type, $F(2, 130)=116$, $p<.001$. Bonferroni-corrected pairwise comparisons indicated that stressed participants ($M=.53$) had lower overall recognition compared to controls ($M=.61$, $p=.03$). Across both groups true recognition ($M=.65$) was better than foil recognition ($M=.35$; $p<.001$) and false recognition of the critical lures ($M=.71$; $p=.02$). In addition, false recognition was significantly higher than foil recognition ($p<.001$), again indicating a robust false recognition effect. Although the Group x Word Type interaction did not reach significance in this analysis, $F(2, 130)=1.98$, $p=.14$, planned comparisons supported our *a*

priori hypothesis that stress would selectively impair true recognition (M=.58, SD=.17 in the stress group vs. M=.71, SD=.13 in the control group), t(65)=3.39; p=.001, while leaving false recognition unaffected (t=.47, p=.64) (Table S1 and Figure 1B. Again, control participants recognized true and false words at similar rates, t(33)=.31, p=.76, but stressed participants recognized proportionately more false words than true words, t(32)=4.23, p<.001. Together with the corrected recognition findings, these results demonstrate that stressed participants not only had impaired veridical or 'true' memory compared to control subjects, but that they were also more likely to falsely recognize critical lure words than they were to correctly recognize studied words.

References

Diekelmann, S., Sandlot, H. P., Lahl, O., Born, J., & Wagner, U. (2008). Sleep loss produces false memories. *PLoS One*, *3*(10).

Feenan, K., & Snodgrass, J. G. (1990). The effect of context on discrimination and bias in recognition memory for pictures and words. Memory & Cognition,18(5), 515-527.

Kirschbaum C, Wust S, Hellhammer D. 1992. Consistent sex differences in cortisol responses to psychological stress. *Psychosom. Med.* 54: 648–657.

Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., & Hellhammer, D. H. (2003). Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change.*Psychoneuroendocrinology*, *28*(7), 916-931.

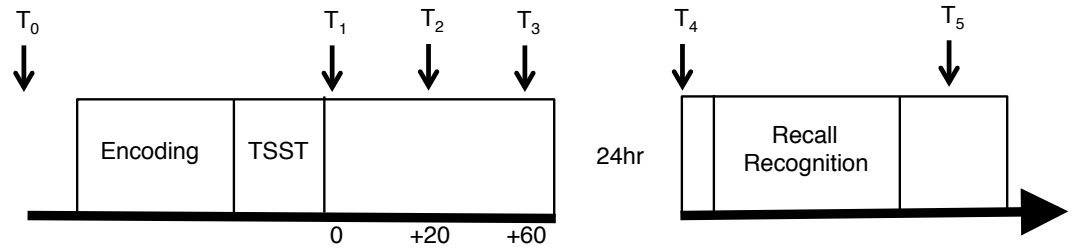
Snodgrass JG, Corwin J. 1988. Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology: General* 117: 34.

Shirtcliff EA, Granger DA, Schwartz E, Curran MJ. 2001. Use of salivary biomarkers in biobehavioral research: cotton-based sample collection methods can interfere with salivary immunoassay results. *Psychoneuroendocrinology* 26: 165-173.

Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & cognition*, *27*(3), 494-500.

Supplementary Figures

Supplementary Figure S1



Supp Fig. S1. Timeline of the experiment. Six saliva samples were taken during the study: at baseline in the afternoon, at 3-4pm (T_0); immediately after the TSST or control task (T_1); 20 min after manipulation (T_2); 60 min after manipulation (T_3); prior to retrieval on the second day (T_4); and at the end of the experiment, prior to leaving the laboratory (T_5).

Supplementary Figure S2

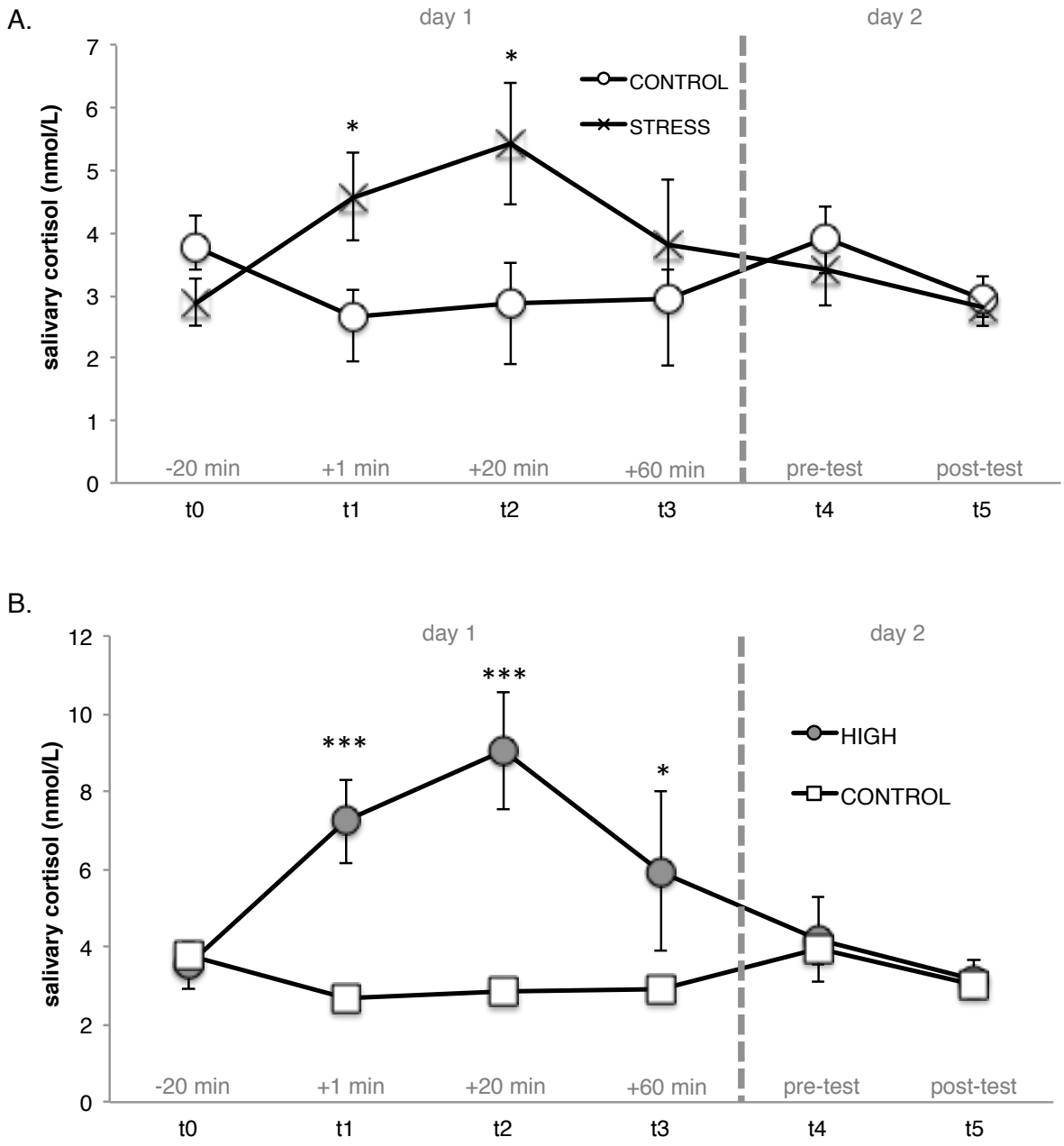
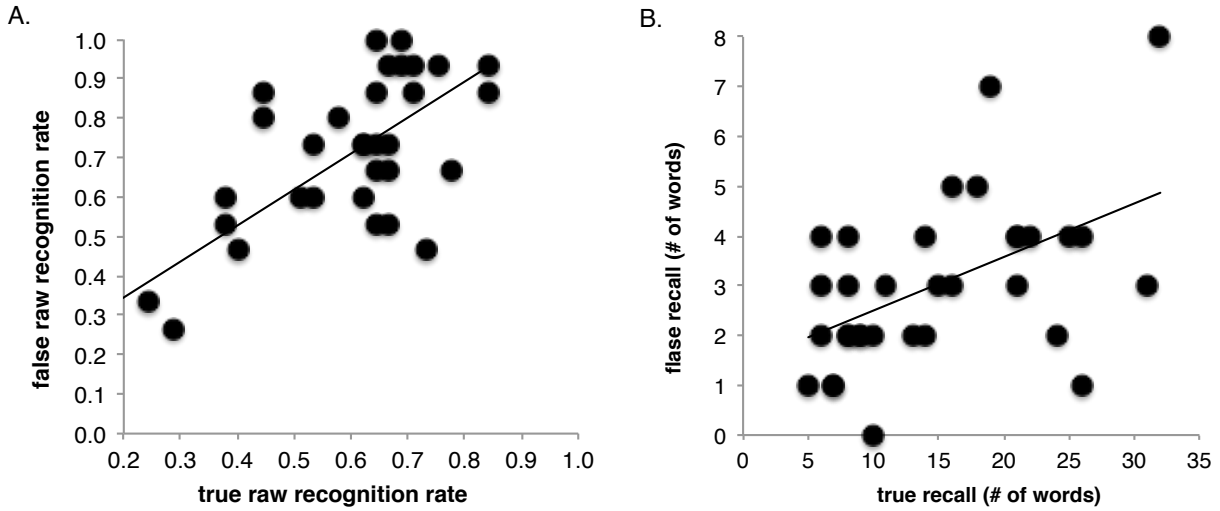


Fig. S2. A) Mean salivary cortisol responses in the control and stress groups (nmol/L). While cortisol levels were equivalent at baseline (t_0), participants in stress group had significantly higher cortisol responses immediately after the TSST (t_1), and 20 minutes after the TSST (t_2). B) Mean salivary cortisol responses in the controls and cortisol HIGH responders. While cortisol levels were equivalent at baseline (t_0), participants in cortisol HIGH responder group had significantly higher cortisol responses immediately after the TSST (t_1), 20 minutes (t_2), and 60 minutes after the TSST (t_3).

* $p<.05$

*** $p<.001$

Supplementary Figure 3



Supp. Fig. S3. Remembering true words is positively correlated to remembering false words, but only in the stress group. These results are observed in both A) recognition rates (r=.72, p<.001) and B) free recall (r=.50, p=.003). This suggests that, for the stress group, the retrieval of study words and critical lures is reliant on similar (gist) processes.

Supplementary Tables

Supplementary Table S1

	Stress (n=33)	Control (n=34)
Raw Recognition		
True rate	.58 (.03)	.71 (.02)
False rate	.70 (.04)	.72 (.04)
Foil rate	.32 (.03)	.39 (.03)
Corrected recognition		
True memory	.27 (.03)	.32(.04)
False memory	.38 (.03)	.33 (.04)
Bias Index Br		
True memory	.43 (.03)	.57 (.03)
False memory	.56 (.05)	.61 (04)

Table S1. Recognition Memory Performance in the Stress and Control Groups. Standard deviations are given in parentheses.

Supplementary Table S2

	Stress (n=33)	Control (n=34)
Free recall		
Studied words	14.9 (1.4)	17.5 (1.7)
Critical lures	3.1 (.29)	3.0 (.29)
Intrusions	6.9 (1.1)	7.2 (1.2)
Free recall rate		
Studied words	.08 (.01)	.09 (.01)
Critical lures	.20 (.02)	.20 (.02)

Table S2. Recall Memory Performance in the Stress and Control Groups. Standard deviations are given in parentheses. The recall rate of studied words is calculated out of a total of 180 study words. The recall rate of critical lures is calculated out of a total of 15 possible lure words.